

1.4. ÍNDICES DE FORMA

En ellos reside una de las principales aportaciones de la estrategia descriptiva EDA, siendo complementada su información por los índices gráficos que trataremos a continuación. Como ya hemos comentado, con distribuciones asimétricas o multimodales los índices clásicos son poco precisos, manifestándose principalmente esta carencia en los índices de forma, que aumentan su robustez mediante los siguientes algoritmos de cálculo:

1.4.1. ÍNDICE DE SIMETRÍA DE YULE

Se calcula mediante:

$$H_1 = \frac{C_{25} + C_{75} - (2 \cdot Md)}{2 \cdot Md} \quad (1.9)$$

que, de nuevo observamos, hace referencia a la simetría del 50% central de la distribución. Su interpretación se realiza como sigue:

- * $H_1 = 0$ → Distribución simétrica
- * $H_1 > 0$ → Asimetría positiva (sesgo o carencia de datos en la mitad superior de la distribución).
- * $H_1 < 0$ → Asimetría negativa (sesgo o carencia de datos en la mitad inferior de la distribución).

1.4.2. ÍNDICE DE SIMETRÍA DE KELLY

Se calcula mediante:

$$H_2 = Md - \frac{C_{10} + C_{90}}{2} \quad (1.10)$$

moors

a quatele abberhe fe kiba

que, como se puede observar, indica la simetría de la distribución en sus extremos o colas. Por ello puede ser de gran utilidad obtener ambos índices, H_1 y H_2 , en base a la complementariedad de su información. Dado que el índice de Kelly depende de las unidades de medida (al emplear la Md), es conveniente transformarlo en el siguiente índice:

$$H_3 = \frac{C_{10} + C_{90} - (2 \cdot Md)}{2 \cdot Md} = \frac{-H_2}{Md} \quad (1.11)$$

interpretándose de esta forma al igual que se hace con el índice de Yule, en base a su adimensionalidad.

1.4.3. COEFICIENTE DE CURTOSIS

Se calcula mediante:

$$K_2 = \frac{C_{90} - C_{10}}{1.9(C_{75} - C_{25})} \quad (1.12)$$

o bien empleando octiles:⁵

$$K_1 = \frac{C_{87.5} - C_{12.5}}{1.7(C_{75} - C_{25})} \quad (1.13)$$

En nuestro ejemplo, el centil 12.5 se corresponde al valor 10 en ambas muestras, en tanto que el centil 87.5 al 18.25 en la muestra 1 y al valor 27.75 en la muestra 2.

Ambos índices se hallan centrados sobre el valor 1, por lo que su interpretación es:

- * K_2 ó $K_1 = 1 \longrightarrow$ Distribución Mesocúrtica
- * K_2 ó $K_1 > 1 \longrightarrow$ Distribución Leptocúrtica
- * K_2 ó $K_1 < 1 \longrightarrow$ Distribución Platicúrtica

5. centiles que dividen a la muestra en 8 partes iguales.

En general, para finalizar la descripción de los datos es estrictamente necesario que se precise la Estadística más conveniente, pues, en caso de no ajustarse, los índices EDA ven

En el pequeño ejemplo siguiendo algoritmos

$\bar{Q}$
TRI
MID
IQR
MAE
PSD₁
PSD₂
CV_c
 H_1
 H_2
 H_3
 K_1
 K_2

Los índices estadísticos, desde un punto de vista ordinario, dan un estudio a fondo de los datos, dado el enfoque empleado. Sin embargo, sí es conveniente que los han propuesto para la diferenciación entre los datos que se realizan en su análisis hasta producirse conv

Presentation Supplement: Proofs of the Selected Theorems

Volkan Cevher
Georgia Institute of Technology
Atlanta, GA
cevher@ieee.org

I. THE FACTORIZATION THEOREM

The factorization theorem is introduced at Slide 15. The proof of this theorem is done for the case in which Γ is discrete and is due to [1]. A general proof can be found in [2].

Let $p_\theta(y|t)$ denote the density of y given $t = T(y)$. By the Bayes formula one have

$$\begin{aligned} p_\theta(y|t) &\triangleq P_\theta(\mathbf{Y} = y | T(\mathbf{Y}) = t) \\ &= \frac{P_\theta(T(\mathbf{Y}) = t | \mathbf{Y} = y) P_\theta(\mathbf{Y} = y)}{P_\theta(T(\mathbf{Y}) = t)} \end{aligned} \quad (\text{I.1})$$

Since $P_\theta(T(\mathbf{Y}) = t | \mathbf{Y} = y) = 1$ if $T(\mathbf{Y}) = t$ and 0 if $T(\mathbf{Y}) \neq t$, and $P_\theta(\mathbf{Y} = y) = p_\theta(y)$, Eq.I.1 becomes

$$p_\theta(y|t) = \begin{cases} p_\theta(y)/P_\theta(T(\mathbf{Y}) = t) & \text{if } T(y) = t, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{I.2})$$

Now $P_\theta(T(\mathbf{Y}) = t) = \sum_{y|T(\mathbf{Y})=t} p_\theta(y)$. To prove the if part of the theorem observe the following

$$\begin{aligned} P_\theta(T(\mathbf{Y}) = t) &= \sum_{y|T(\mathbf{Y})=t} g_\theta[T(y)]h(y) \\ &= g_\theta(t) \sum_{y|T(\mathbf{Y})=t} h(y) \end{aligned} \quad (\text{I.3})$$

in addition one also have $p_\theta(y) = g_\theta[T(y)]h(y) = g_\theta(t)h(y)$. From Eq. I.2 one then have

$$p_\theta(y|t) = \begin{cases} h(y) / \sum_{y|T(\mathbf{Y})=t} h(y) & \text{if } T(y) = t, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{I.4})$$

Since the right hand side of Eq. I.4 does not depend on θ , T is a sufficient statistic for the parameter set $\theta \in \Lambda$.

To prove the only if statement in the theorem, let T be any sufficient statistic for θ . From Eq. I.2 one can write

$$p_\theta(y) = p_\theta(y|T(y))P_\theta(T(\mathbf{Y}) = T(y)) \quad (\text{I.5})$$

Since T is sufficient for θ , $p_\theta(y|T(y))$ depends only on y and not on θ . On defining $h(y) \triangleq p_\theta(y|T(y))$ and $g_\theta[T(y)] \triangleq P_\theta(T(\mathbf{Y}) = T(y))$, one can see that Eq. I.5 implies the factorization theorem. Hence, the proof is complete.

II. THE RAO-BLACKWELL THEOREM

Slide 17 presents the Rao-Blackwell theorem, which is very useful for minimum variance unbiased estimators. The

theorem and its proof can also be found in [1].

To prove that $\tilde{g}[T(\mathbf{Y})]$ is unbiased, take the expectation

$$\begin{aligned} E_\theta\{\tilde{g}[T(\mathbf{Y})]\} &= E_\theta\{E_\theta\{\hat{g}(\mathbf{Y})|T(\mathbf{Y})\}\} \\ &\Rightarrow \tilde{g}[T(\mathbf{Y})] = E_\theta\{\hat{g}(\mathbf{Y})\} = g(\theta) \end{aligned} \quad (\text{II.1})$$

First note that the expectation defining $\tilde{g}$ does not depend on θ due to the sufficiency of T . Secondly, the second equality can be obtained by using the fact that $E\{E\{X|Z\}\} = E\{X\}$ and the unbiasedness of $\hat{g}$.

In order to see that $\text{Var}_\theta(\tilde{g}[T(\mathbf{Y})]) \leq \text{Var}_\theta(\hat{g}(\mathbf{Y}))$, note the following

$$\begin{aligned} \text{Var}_\theta(\tilde{g}[T(\mathbf{Y})]) &= E_\theta\{[\tilde{g}[T(\mathbf{Y})]]^2\} - g^2(\theta) \\ \text{Var}_\theta(\hat{g}(\mathbf{Y})) &= E_\theta\{[\hat{g}(\mathbf{Y})]^2\} - g^2(\theta) \end{aligned} \quad (\text{II.2})$$

Hence, if it can be shown that $E_\theta\{[\tilde{g}[T(\mathbf{Y})]]^2\} \leq E_\theta\{[\hat{g}(\mathbf{Y})]^2\}$, the proof is complete.

$$\begin{aligned} E_\theta\{[\tilde{g}[T(\mathbf{Y})]]^2\} &= E_\theta\{[E_\theta\{\hat{g}(\mathbf{Y})|T(\mathbf{Y})\}]^2\} \\ &\leq E_\theta\{E_\theta\{[\hat{g}(\mathbf{Y})]^2|T(\mathbf{Y})\}\} \\ &= E_\theta\{[\hat{g}(\mathbf{Y})]^2\}, \end{aligned} \quad (\text{II.3})$$

The second equality follows from Jensen's inequality¹ and the final equality follows from iterated expectation operations. The equality in Jensen's inequality is satisfied if and only if $P_\theta[\hat{g}(\mathbf{Y}) = E_\theta\{\hat{g}(\mathbf{Y})|T(\mathbf{Y})\} | T(\mathbf{Y})] = 1$, and using the definition of $\tilde{g}$ it is easy to see that this condition is equivalent to $P_\theta[\hat{g}(\mathbf{Y}) = \tilde{g}[T(\mathbf{Y})]] = 1$. This completes the proof of the Rao-Blackwell theorem.

III. CRAMER-RAO BOUND

The Cramer-Rao bound establishes a lower bound on the error covariance matrix for any unbiased estimator $\hat{\theta}$ for a parameter θ and was introduced in Slide 39. To set up the Cramer-Rao bound, we need to define a function called the score function, interpret it, and establish its statistical properties. The proof here follows the one in chapter 6 of [3].

The score function is defined to be the gradient of the log-likelihood function:

¹*Jensen's Inequality*: For any random variable X and convex function C , $E\{C(X)\} \geq C(E\{X\})$ with equality if and only if $P(X = E\{X\}) = 1$ when C is strictly convex.

$$s(\theta, \mathbf{y}) = \frac{\partial}{\partial \theta} L(\theta, \mathbf{y}) = \frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{y}) \quad (\text{III.1})$$

When the realization $\mathbf{y}$ is replaced by the random variable $\mathbf{Y}$, then the log-likelihood and score functions become random variables:

$$s(\theta, \mathbf{Y}) = \frac{\partial}{\partial \theta} L(\theta, \mathbf{Y}) = \frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{Y}) \quad (\text{III.2})$$

The score function scores values of θ as the random vector $\mathbf{Y}$ assumes values from the distribution $p_{\theta}(\mathbf{y})$. Scores are good scores and scores different from zero are bad scores. The score function has zero mean:

$$\begin{aligned} E\{s(\theta, \mathbf{Y})\} &= E\left\{\frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{Y})\right\} \\ &= \int d\mathbf{y} p_{\theta}(\mathbf{y}) \frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{y}) \\ &= \int d\mathbf{y} \frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{y}) = \frac{\partial}{\partial \theta} \int d\mathbf{y} p_{\theta}(\mathbf{y}) = \mathbf{0} \end{aligned} \quad (\text{III.3})$$

The covariance matrix of the score function $s(\theta, \mathbf{Y})$ is called the *Fisher information matrix* and is denoted by $\mathbf{J}(\theta)$:

$$\mathbf{J}(\theta) = E\{s(\theta, \mathbf{Y})s^T(\theta, \mathbf{Y})\} = E\left\{\frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{Y})\left(\frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{Y})\right)^T\right\}. \quad (\text{III.4})$$

This result for the Fisher information matrix can be cast in a different, but equivalent, form by noting that the function $\frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{y})$ may be rewritten as

$$\frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{y}) = \frac{1}{p_{\theta}(\mathbf{y})} \frac{\partial}{\partial \theta} p_{\theta}(\mathbf{y}). \quad (\text{III.5})$$

The second gradient of $\log p_{\theta}(\mathbf{y})$ may then be rewritten as

$$\frac{\partial}{\partial \theta} \left(\frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{y}) \right)^T = \frac{\frac{\partial}{\partial \theta} \left(\frac{\partial}{\partial \theta} p_{\theta}(\mathbf{y}) \right)^T}{p_{\theta}(\mathbf{y})} - \frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{y}) \left(\frac{\partial}{\partial \theta} p_{\theta}(\mathbf{y}) \right)^T. \quad (\text{III.6})$$

The expectation of the first term on the right-hand side is zero, so

$$E\left\{\frac{\partial}{\partial \theta} \left(\frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{Y}) \right)^T\right\} = -E\left\{\frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{Y}) \left(\frac{\partial}{\partial \theta} p_{\theta}(\mathbf{Y}) \right)^T\right\}. \quad (\text{III.7})$$

This identity produces formula for the Fisher information matrix:

$$\mathbf{J}(\theta) = -E\left\{\frac{\partial}{\partial \theta} \left(\frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{Y}) \right)^T\right\}. \quad (\text{III.8})$$

These results are summarized by recording the i, j element of the Fisher information matrix:

$$\begin{aligned} [\mathbf{J}(\theta)]_{i,j} &= E\left\{\frac{\partial}{\partial \theta_i} \log p_{\theta}(\mathbf{Y}) \left(\frac{\partial}{\partial \theta_j} \log p_{\theta}(\mathbf{Y}) \right)^T\right\} \\ &= E\left\{\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p_{\theta}(\mathbf{Y})\right\} \end{aligned} \quad (\text{III.9})$$

There is one more property we will need. The cross-covariance between the score function and the error of any unbiased estimator $\hat{\theta}$ is identity:

$$E\{s(\theta, \mathbf{Y})[\hat{\theta} - \theta]^T\} = \mathbf{I} \quad (\text{III.10})$$

To establish this remarkable property, we note that the unbiasedness of $\hat{\theta}$ implies $E\{[\hat{\theta} - \theta]^T\} = \mathbf{0}^T$. This may be written as $\int d\mathbf{y} p_{\theta}(\mathbf{y})[\hat{\theta} - \theta]^T = \mathbf{0}^T$. Taking the gradient with respect to θ , one can obtain:

$$\begin{aligned} \int d\mathbf{y} \frac{\partial}{\partial \theta} p_{\theta}(\mathbf{y})[\hat{\theta} - \theta]^T - \int d\mathbf{y} p_{\theta}(\mathbf{y}) \mathbf{I} &= \mathbf{0} \\ \int d\mathbf{y} p_{\theta}(\mathbf{y}) \frac{\partial}{\partial \theta} \log p_{\theta}(\mathbf{y})[\hat{\theta} - \theta]^T &= \mathbf{I} \end{aligned} \quad (\text{III.11})$$

$$E\{s(\theta, \mathbf{Y})[\hat{\theta} - \theta]^T\} = \mathbf{I}.$$

Then the error covariance matrix for $\hat{\theta}$ is bounded as follows:

$$\mathbf{C} = E\{[\hat{\theta} - \theta][\hat{\theta} - \theta]^T\} \geq \mathbf{J}^{-1}, \quad (\text{III.12})$$

provided that $\mathbf{J}$ is positive definite. That is, the matrix $\mathbf{C} - \mathbf{J}^{-1}$ is nonnegative definite, as is the matrix $\mathbf{J} - \mathbf{C}^{-1}$.

$$\begin{bmatrix} \hat{\theta} - \theta \\ s(\theta, \mathbf{Y}) \end{bmatrix} \quad (\text{III.13})$$

This vector has zero mean. Its covariance matrix is given by

$$\begin{aligned} \mathbf{Q} &= E\left\{\begin{bmatrix} \hat{\theta} - \theta \\ s(\theta, \mathbf{Y}) \end{bmatrix} \begin{bmatrix} \hat{\theta} - \theta \\ s(\theta, \mathbf{Y}) \end{bmatrix}^T\right\} \\ &= \begin{bmatrix} \mathbf{C} & \mathbf{I} \\ \mathbf{I} & \mathbf{J} \end{bmatrix} \end{aligned} \quad (\text{III.14})$$

The nonnegative definite covariance matrix $\mathbf{Q}$ may be diagonalized as follows:

$$\begin{bmatrix} \mathbf{I} & -\mathbf{J}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{C} & \mathbf{I} \\ \mathbf{I} & \mathbf{J} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{J}^{-1} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{C} - \mathbf{J}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} \end{bmatrix} \quad (\text{III.15})$$

Thus, the covariance matrix $\mathbf{Q}$ is similar to the matrix on the right-hand side. Therefore, $\mathbf{C} - \mathbf{J}^{-1}$ is nonnegative definite, meaning $\mathbf{C} \geq \mathbf{J}^{-1}$ or $\mathbf{J} \geq \mathbf{C}^{-1}$. The i, i element of $\mathbf{C}$ is the mean-squared error of the estimator of θ_i :

$$C_{i,i} = E\{(\hat{\theta}_i - \theta_i)^2\} \geq (\mathbf{J}^{-1})_{i,i}. \quad (\text{III.16})$$

So, the i, i element of the inverse of the Fisher information matrix lower bounds the mean-squared error of any unbiased estimator of θ_i .

REFERENCES

- [1] Poor, H. V. (1994), *An Introduction to Signal Detection and Estimation* (Downs & Culver, Inc.)
- [2] Lehmann, E. L. (1986), *Testing Statistical Hypotheses* (Wiley: New York)
- [3] Scharf, L. S. (1991), *Statistical Signal Processing* (Addison-Wesley Publishing Company, Inc.)